



Modelling N₂O Emissions in Fertilized Grasslands with Machine Learning: A Multi-Site LSTM Approach

Samin Payrosangari¹, Kathrin Fuchs¹, Benjamin Wolf¹, Ralf Kiese¹, Clemens Scheer¹

¹Karlsruhe Institute of Technology, IMK-IFU, Garmisch-Partenkirchen, Germany

Correspondence to: Samin Payrosangari, Kathrin Fuchs (samin.payrosangari@kit.edu, kathrin.fuchs@kit.edu)

Abstract. Accurate quantification of GHG emissions from agricultural soils based on key environmental and management factors, is crucial for developing effective mitigation strategies. This study applies a machine learning approach using data from multiple montane grasslands in central Europe to model the spatial and temporal dynamics of N₂O emissions, using meteorological, soil, and management data. The primary aim is to advance predictive modelling of N₂O emissions from grassland soils for spatial upscaling and scenario analysis. Specifically, we assess whether a generic model can accurately estimate N₂O emissions from an independent site excluded from training.

We collected data from five fertilized grasslands across southern Germany, northern Switzerland and northern Austria (122 soil-site-treatment-year combinations), and trained a long short-term memory (LSTM) algorithm to model the influence of drivers within 5 subsequent days on N₂O emissions. The dataset includes daily N₂O emission measurements along with key emission drivers such as soil moisture and temperature in 10 cm soil depth, daily precipitation, occurrence of fertilization events (zero or one flag), as well as soil characteristics such as pH and bulk density.

The trained LSTM model showed strong predictive performance (RMSE of 18 $\mu\text{gm}^{-2}\text{h}^{-1}$, and Relative RMSE of 270%) when evaluated on a test set that included both data from an independent soil and withheld years from training procedures. The model accurately captured N₂O dynamics, including the magnitude and timing of emission peaks driven by slurry application and environmental factors. Compared to the performance of established process-based biogeochemical models the LSTM model yielded similar RMSE and bias values for most site-years. These results demonstrate that LSTM-based models can reliably predict N₂O emissions at independent sites with similar environmental and soil characteristics and represent a promising alternative to process-based models for predicting soil N₂O emissions.

1 Introduction

Greenhouse gas (GHG) emissions significantly contribute to global warming, with agricultural activities accounting for approximately 12% of global anthropogenic GHG emissions (Chataut et al., 2023). The greenhouse gases emitted from soils are mainly methane (CH₄), nitrous oxide (N₂O) and carbon dioxide (CO₂), which are increasing due to high-energy-intensive agricultural production systems and the production and use of mineral and organic fertilizers (Chataut et al., 2023). Today, agricultural activities are responsible for approx. two-thirds of total anthropogenic N₂O emissions and its high global warming potential makes it the third-most important greenhouse gas for global warming (Schulz et al., 2021; Fuchs et al., 2020). Thus, agricultural mitigation strategies play an important role in reducing N₂O emissions, which requires understanding the effects of management adoption (Fuchs et al., 2018).

The production of N₂O in the soil is driven by microbial processes — predominately denitrification and nitrification — which are influenced by both environmental conditions and management practices (Wang et al., 2021; Butterbach-Bahl et al., 2013). Key soil and environmental conditions regulating these processes include soil moisture, temperature, texture, nitrogen content, pH, and organic matter content, while fertilization, tillage, and irrigation represent the most important management practices (Wang et al., 2021; Butterbach-Bahl et al., 2013). The Pearson correlation between daily N₂O emissions and individual input drivers might be low (between -0.11 to 0.16 for our dataset), and it's the interaction of drivers such as precipitation and temperature, as well as management practices that result in highly dynamic, variable and site-specific emission patterns (Wang et al., 2021). Thus, the complex influences of different drivers on emissions need to be simulated via high-performance models. This is especially relevant as direct measurement over large areas and long periods is impractical (Butterbach-bahl et al., 2013; Giltrap et al., 2010).

Modelling approaches for estimating N₂O emissions from agricultural systems include emission factors, statistical models, and process-based models (Butterbach-Bahl et al., 2013). Emission factors, i.e., a fixed proportion of applied nitrogen that is emitted as N₂O, are widely used for national inventory

reporting. However, they are unable to reflect the variability of soil and climate conditions that are relevant for N₂O emissions at the field scale or with a time resolution higher than yearly (Giltrap et al., 2010). In contrast, process-based models such as LandscapeDNDC (Haas et al. 2013) or Daycent (Parton et al., 1998) simulate the interactions between soil processes, climatic drivers and management. To this end, these models use a large set of parameters, which need to be parameterized based on calibration and validation datasets. As a consequence, uncertainty of model results due to parameterization is inevitable and the performance of such models heavily depends on the number and quality of calibration and validation datasets.

Given these limitations, machine learning (ML) offers a promising alternative or complementary approach to process-based modelling. ML models are trained on previous measurements and do not require a deep understanding or a definition of the underlying processes and parameters. However, implementing a high-accuracy ML model comes with challenges, such as the need for large, high-quality datasets and expert knowledge about the used algorithms, along with lower interpretability than process-based models. While implementing the right ML model requires effort, once a generic model is built, applying it to new data is highly time-efficient.

Building an ML model for estimating N₂O emissions requires considering short-term fluctuations and peaks in N₂O emissions in response to temporal dynamics in soil moisture and temperature, such as dry–wet transitions, water logging, as well as management activities (Butterbach-Bahl et al., 2013). Therefore, estimating daily values of N₂O emissions can be improved by considering the changes in drivers over several days prior to the emission event, which is addressed in this study. Long short-term memory network or LSTM (Hochreiter and Schmidhuber, 1997) is a machine learning model that detects the dynamic relationships in sequential data. It has previously been used for different applications such as flood forecasting (Le et al., 2019), stock market prediction (Chhajjer et al., 2022), and predicting water table depth in agricultural areas (Zhang et al., 2018), and therefore is a promising method for modelling soil N₂O emissions.

In this study, we developed an LSTM model to simulate the daily N₂O emissions of grasslands in central Europe, with the three main objectives:

- 1) to train an LSTM model using data from multiple soils and climates to simulate daily N₂O emissions at alpine grasslands
- 2) to test the model performance on an independent grassland soil not used for model training to assess the general applicability of the model to alpine grasslands without site-specific parametrization
- 3) to assess the relative importance of different environmental and management variables for predicting N₂O emissions

2 Materials and methods

Daily and growing season N₂O emission data were collected from multiyear field experiments of 5 grassland sites representing a variety of soils at different elevations and thus climate conditions in Central Europe: South Germany (Fendt, Rottenbuch and Graswang), North Austria (Neustift) and North Switzerland (Chamau). Table 1 shows the characteristics of these grasslands, all of which are managed with organic fertilizer (manure and slurry), cutting and no grazing. While N₂O emissions were highest at Chamau, lowest N₂O emissions were observed at the Neustift site. The collected dataset for training the LSTM model includes high-resolution daily measurements of soil N₂O emissions, soil physical and chemical properties (Table 2), as well as daily meteorology

Grassland	Elevation (m asl)	Annual N ₂ O (kg h ⁻¹ y ⁻¹)	Annual precipitation (mm)	Avg annual air temperature (°C)	pH	BD (gr cm ⁻³)	Dataset size
Chamau Switzerland (CH-CHA)	393	5.28	1179	9.8	6.5	0.94	1740
Fendt Germany (DE-Fen)	595	2.83	962	9.3	4.8-6.5	0.8-1.1	8195
Rottenbuch Germany (DE-RbW)	769	1.69	1047	8.6	4.8-5.7	0.7-0.95	8690
Graswang Germany (DE-Gwg)	864	2.55	1464	7	5.1-6.6	0.52-0.78	8708
Neustift Austria (AT-NEU)	970	0.03	852	6.5	5.7	1.2	259

Table 1 Summary of the properties of grasslands used in this study (Kiese et al., 2018; Soltani et al., 2018; Gilgen & Buchmann, 2009; Roth, 2006; Hörtnagel et al., 2014). The total annual N₂O is averaged over different years of data availability for each grassland.



2.1 Site descriptions

70 German Terrestrial Environmental Observatories (TERENO) Pre-Alpine

The Tereno Pre-Alpine observatory includes three grassland sites, Graswang (DE-Gwg), Rottenbuch (DE-RbW), and Fendt (DE-Fen), at elevations of 864, 769, and 595 m asl, which are located in the Bavarian Prealps in southern Germany (Kiese et al., 2018; Soltani et al., 2018). Automatic chamber measurements of N₂O emission along with a wide range of other measurements such as soil moisture and temperature are conducted by the Karlsruhe Institute of Technology (KIT). The automatic chamber measurements are carried out on 36 lysimeters, i.e., soil cores with 1.0 m² surface and a depth of 1.4 m (Kiese et al., 2018). A subset of 18 out of 36 soil cores are translocated from the high-elevation site Graswang and the medium-elevation site Rottenbuch to the lower altitude sites (space-for-time-approach), referred to as Graswang in Fendt, Graswang in Rottenbuch or Rottenbuch in Fendt translocations to investigate the soils under climate change scenarios (Kiese et al., 2018). The lysimeter setup for these grasslands is shown in Fig. A1 of appendix A. Experiments on each lysimeter follow intensive or extensive management treatments, where extensive management includes up to three cuts and two manure applications, and intensive management includes up to six cuts and five slurry applications per year (Kiese et al., 2018). From each soil and treatment three replicates are considered, for instance, there are three Graswang in Fendt soil core replicates which are intensively managed as shown in Fig. A1. Notably, at the time of this study, sub-daily (up to 6 flux rates per day) data from 24 of the lysimeters being all lysimeters operated in Fendt (Fig. A1 a) and Graswang (Fig. A1 c), which are aggregated into daily are used according to availability of data for different years between 2014 and 2020. The soil properties of different replicates (N=3: L1-L3) are shown in Table 2.

Switzerland FLUXNET Chamau

85 Chamau (CH-CHA) is a permanent grassland site that was established by ETH Zurich in 2005 (Feigenwinter et al., 2023) and is part of the FLUXNET initiative. This site is located in the Reuss River valley on the Swiss Plateau, approximately 30 km south of the city of Zurich, Switzerland, at an elevation of 393 m asl. Half-hourly eddy-covariance measurements of N₂O and CH₄ started at this site in 2012 (Merbold et al., 2014). Moreover, a N₂O mitigation experiment, from 2015 to 2020, was conducted by separating two parcels at north and south of the site, continuing slurry fertilization on the control parcel in south and oversowing the north parcel with a clover seed mixture without fertilization (Fuchs et al., 2018). The measurements from the control parcel 90 aggregated into daily values are used in this study, which represent an intensively managed site with cutting 4–5 times a year followed by liquid slurry application plus one slurry application in spring each year (Feigenwinter et al., 2023; Hörtnagl et al., 2025).

Austria Neustift

The Neustift grassland (AT-NEU) is an intensively managed meadow in the middle of a flat valley at the bottom of the Stubai valley in the Austrian alps, at an elevation of 970 m a.s.l (Hörtnagl & Wohlfahrt, 2014). The N₂O and CH₄ emissions at this site were measured between April 2010 and February 2012, 95 and the management consisted of three cuts and one solid manure and slurry application in autumn each year, which was considered an intensively managed site (Hörtnagl & Wohlfahrt, 2014; Hörtnagl et al., 2018). At this site, fluxes were calculated via virtual disjunct eddy-covariance, providing half-hourly flux measurements (Hörtnagl & Wohlfahrt, 2014), which are aggregated to daily values for this study.



Soil	Climate (Grassland)	Replicate-Management	Clay 0-10 cm (%)	Silt 0-10 cm (%)	Sand 0-10 cm (%)	SOC 0-10 cm (%)	Nt 0-10 cm (%)	C/N 0-10 cm	pH 0-10 cm	Bulk density 0-10 cm (gr cm ⁻³)	Dataset assignment	Data availability years	Dataset size
Graswang	Graswang	L1-Intensive	55	42	3	11.4	1.2	9.48	6.40	0.52	Train	2020	277
Graswang		L2-Intensive	38	43	19	8.7	0.81	10.74	5.43	0.73	Train	2020	273
Graswang		L3-Intensive	59	35	6	9.5	0.83	9.53	6.02	0.68	Train	2020	269
Graswang	Fendt	L1-Intensive	61	36	3	10.9	1.15	9.43	6.56	0.55	Train	2014-2020	908
Graswang		L2-Intensive	37	44	19	7.5	0.73	10.27	5.35	0.78	Train	2014-2020	1190
Graswang		L3-Intensive	61	35	4	8.5	0.9	9.44	6.12	0.66	Train	2014-2020	1239
Graswang		L1-Extensive	57	40	3	11.3	1.195	9.42	6.40	0.52	Train	2014-2020	1327
Graswang		L2-Extensive	58	38	5	7.7	0.81	9.51	6.12	0.53	Train	2014-2020	1232
Graswang		L3-Extensive	38	43	20	8.0	0.82	9.76	5.10	0.77	Train	2014-2020	1142
Fendt		L1-Intensive	25	40	35	5.2	0.52	9.92	5.87	0.93	Train	2014-2020	1216
Fendt		L2-Intensive	30	43	27	4.5	0.45	9.89	4.87	1.01	Train	2014-2020	1129
Fendt		L3-Intensive	33	46	21	6.4	0.68	9.35	5.16	0.78	Train	2014-2020	1379
Chamau	Chamau	Intensive	19	45	36	3.2	0.33	9.79	6.50	0.94	Train/Test	2013, 2014, 2016, 2019, 2020/ 2015	1420/ 320
Neustift	Neustift	Intensive	4	43	53	3.0	0.28	10.71	5.70	1.20	Train	2010-2011	259
Rottenbuch	Fendt	L1-Extensive(2014-2018), Intensive(2019-2020)	40	41	19	7.0	0.725	9.66	5.73	0.69	Test	2014-2020	1320
Rottenbuch		L2-Extensive(2014-2018), Intensive(2019-2020)	24	42	34	4.3	0.44	9.75	5.34	0.95	Test	2014-2020	1452
Rottenbuch		L3-Extensive(2014-2018), Intensive(2019-2020)	23	53	24	5.8	0.63	9.19	5.43	0.81	Test	2014-2020	1478
Rottenbuch		L1-Intensive	24	43	33	4.4	0.47	9.32	4.84	0.93	Test	2014-2020	1554
Rottenbuch		L2-Intensive	36	40	24	6.2	0.63	9.84	5.39	0.70	Test	2014-2021	1532
Rottenbuch		L3-Intensive	28	49	23	5.9	0.65	9.08	5.18	0.80	Test	2014-2022	1354
Fendt		L1-Extensive(2014-2018), Intensive(2019-2020)	27	42	30	5.5	0.57	9.72	5.72	0.86	Test	2014-2020	1755
Fendt		L2-Extensive(2014-2018), Intensive(2019-2020)	31	42	28	4.4	0.45	9.69	4.90	0.96	Test	2014-2020	1399
Fendt		L3-Extensive(2014-2018), Intensive(2019-2020)	34	45	21	6.4	0.69	9.32	5.26	0.81	Test	2014-2020	1317
Graswang	Graswang	L1-Extensive	60	35	5	8.5	0.87	9.72	6.07	0.65	Test	2020	284
Graswang		L2-Extensive	58	39	0	11.1	1.13	9.86	6.45	0.55	Test	2020	284
Graswang		L3-Extensive	37	44	19	8.1	0.75	10.80	5.20	0.80	Test	2020	283

Table 2 Soil properties of the different soils in various climates used in this study (Kiese et al., 2018; Soltani et al., 2018; Gilgen & Buchmann, 2009; Roth, 2006; Hörtnagel et al., 2014). All soil properties refer to measurements in 0-10 cm depth.

2.2 Input variables and model setup

In this study we model the average daily N₂O emissions based on the temporal and static drivers available in the collected dataset. The temporal drivers used by the model are the daily averages of soil moisture and temperature at 10 cm soil depth, total precipitation, and the occurrence of a fertilization event. The model also incorporates static inputs, such as soil pH and bulk density, to represent the soil's chemical and physical properties across the different sites. Each individual data point of the problem can be expressed as Eq. 1:

$$\left\{ \begin{array}{l} \text{Soil mois}_{t-4} \\ \text{Soil mois}_{t-3} \\ \text{Soil mois}_{t-2} \\ \text{Soil mois}_{t-1} \\ \text{Soil mois}_t \end{array} \right\}, \left\{ \begin{array}{l} \text{Soil temp}_{t-4} \\ \text{Soil temp}_{t-3} \\ \text{Soil temp}_{t-2} \\ \text{Soil temp}_{t-1} \\ \text{Soil temp}_t \end{array} \right\}, \left\{ \begin{array}{l} \text{Precipitation}_{t-4} \\ \text{Precipitation}_{t-3} \\ \text{Precipitation}_{t-2} \\ \text{Precipitation}_{t-1} \\ \text{Precipitation}_t \end{array} \right\}, \left\{ \begin{array}{l} \text{Fertilization}_{t-4} \\ \text{Fertilization}_{t-3} \\ \text{Fertilization}_{t-2} \\ \text{Fertilization}_{t-1} \\ \text{Fertilization}_t \end{array} \right\}, \{pH, BD\} \rightarrow \{N_2O_t\} \quad (1)$$

↓ Temporal
 ↓ Static
 ↓ Target

The indices indicate the values of the features at different days in which day t is the current day. The average daily N₂O emissions at day t is the target of the machine learning model.

2.3 Dataset structure and pre-processing

The availability of a diverse dataset that includes translocated soil cores among German grassland, also using data from Switzerland and Austria gives this study the opportunity to cover a variety of soils in different climates (Table 2), which is key in creating a generic model capable to transfer results across sites. At German sites, replicates sampled from individual grassland often exhibit varying soil properties, as they are taken from different sections of expansive grassland areas. Therefore, for training the model the data from each replicated lysimeter was used separately so that the model could learn the role of soil properties in more detail. Splitting the dataset to training and test sets is not done by random shuffling, but according to the objective of the study a different approach is taken. We split the dataset based on year and management technique and completely exclude the Rottenbuch soils from the model development process to consider it as an independent soil to assess the transferability of the model. For the training set, the following prerequisites are considered: 1) include treatments with a variety of soil properties, N₂O emission patterns and climates; 2) include soils with intensive treatment to ensure that the model has enough data points with fertilization events to learn the role of fertilizer under different conditions; and 3) the data from Neustift was only used for training, as a grassland with highest altitude and lowest emissions. The training set includes around 13000 data sequences in total.



On the other hand, for the test set, 3 important criteria are considered according to the goals of the study: 1) Firstly, from the Chamau soil some years are used for training, and 1 year data used as a test set to show the ability of the model in gap-filling this site for unseen years; 2) secondly, from Fendt and Graswang soils, some of the lysimeters with extensive management are used for testing, 3) Finally, all the available data from the Rottenbuch site is held back for use as an independent test soil. The test set includes around 15000 data sequence.

Along with properly splitting the dataset, data cleansing and pre-processing is crucial for training the ML model. In this study the quality checked measured data from which the gaps were excluded are used. In addition, scaling the data prior to training ensures faster convergence and improves model performance by preventing features with larger magnitudes from disproportionately influencing the learning process. The input drivers are all scaled via the Scikit-learn library robust scaler (Pedregosa et al., 2011) shown in Eq. 1. The choice of using robust scaler is made after investigating other options such as standardization and normalization (min-max scaler) for all or individual drivers according to their distribution. The robust scaler transforms each feature by removing the median and scaling it according to the interquartile range, making it robust to outliers. For each variable x , the scaled value x_{scaled} is computed via Eq. 2:

$$x_{scaled} = \frac{x - \text{Median}(x)}{Q_3(x) - Q_1(x)} \quad (2)$$

Moreover, the target value (daily N_2O emission) is log transformed via natural logarithm ($\ln$) for model setup, as its original values show a right-skewed distribution as shown in Fig. 1 for the training set. Therefore, the output of the model is generated at log scale and is subsequently needs to be back transformed to the actual emission scale via the exponential function.

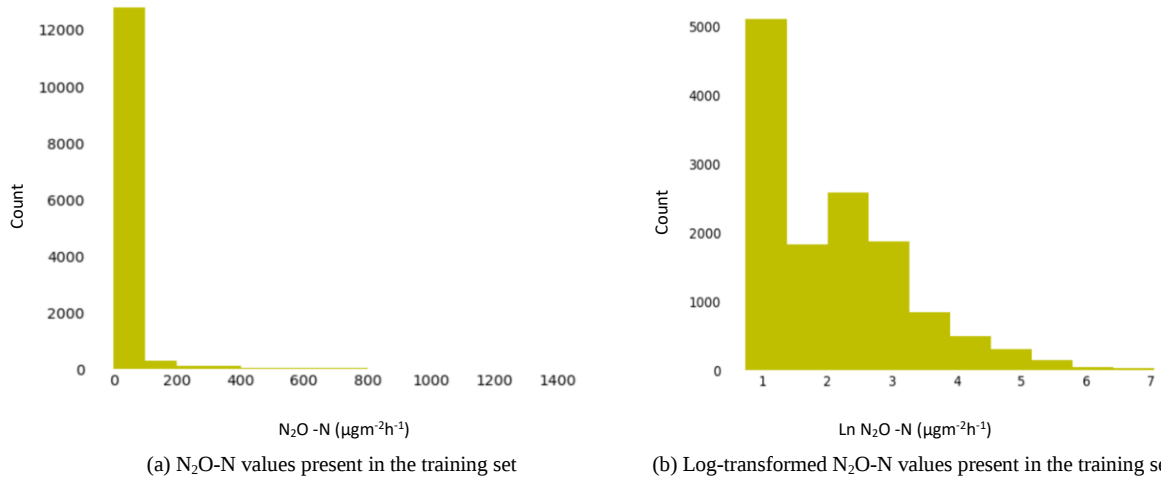


Figure 1 Histogram of (a) original and (b) log-transformed daily $\text{N}_2\text{O-N}$ emissions values at the training set used in this study. Max N_2O emission value equals to $1155 \mu\text{gm}^{-2}\text{h}^{-1}$, min value equals to $0.09 \mu\text{gm}^{-2}\text{h}^{-1}$ and average equals to $17.5 \mu\text{gm}^{-2}\text{h}^{-1}$ among the 13000 training datapoints.



2.4 Model development and training procedure

To build a model based on LSTM neural networks with high performance in simulating daily N₂O emissions, different aspects such as the architecture of the neural network, scaling and transforming the data, selecting a proper loss function for training and hyperparameter tuning should be considered. In this section, the methods used in this study for model development are explained.

2.4.1 LSTM model architecture

Long short-term memory (LSTM) is a type of recurrent neural networks (RNN) that learns long-term dependencies in sequential data and is often used in speech recognition and time series processing (Hochreiter and Schmidhuber, 1997). As discussed in previous sections, simulation of daily N₂O emissions from soil is more accurate if the role and changes in environmental conditions over several days are simulated using a sequential model. In this study, two different network architectures shown in Fig. 2 were implemented using the Keras library in Python (Chollet et al., 2015). Each architecture is investigated further with various configurations and hyperparameters to find the optimum model.

The first architecture, Model A, firstly processed the temporal inputs (Eq. 1) via a set of LSTM layers. Subsequently, several dense layers that receive the output of the LSTM layers as well as a set of static inputs (pH and bulk density) are added to be processed by the model.

In contrast, the second architecture, Model B, initially processes the static inputs via dense layers, then uses the processed information as an initial input to a set of LSTM layer (initializing the hidden state of first LSTM layer) and includes the temporal inputs at this stage.

The models in this study are built with an attention mechanism. Attention in LSTM refers to a mechanism that allows the model to dynamically focus on specific parts of the input sequence when generating each output, improving performance by weighting relevant hidden states (Bahdanau et al., 2015).

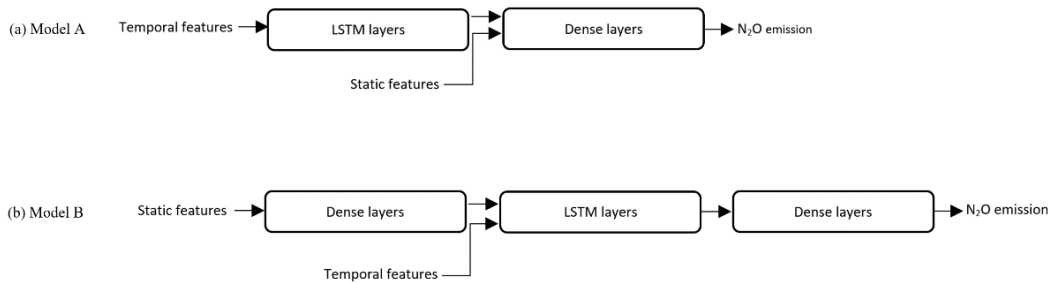


Figure 2 Two basic architectures of the LSTM tested in this study.

2.4.2 Loss function

The loss function used during training quantifies the error between the model's predictions and the actual target values, guiding the optimization process to minimize this error. For the current imbalanced dataset, a custom loss function is used to highlight the importance of catching high peaks of N₂O emissions which is formulated in Eqs. (3) and (4) as:

$$loss = abs(y_{true} - y_{pred}), \quad (3)$$

$$weighted\ loss = loss * e^{loss_coefficient * loss},$$

(4)



With y_{true} and y_{pred} representing the observed and predicted daily N_2O fluxes, respectively. This loss function is significantly penalizing the model for high errors in peak values and showed a better performance comparing to the mean squared error (MSE), mean absolute error (MAE) as well as weighted mean squared (WMSE) error. In addition, loss_coefficient is a hyperparameter controlling the exponential weight further, with potential values shown in Table 3.

190 2.4.3 Hyperparameter tuning

Hyperparameter tuning is a key step in developing machine learning models, as it helps identifying the best-performing configuration for a given problem. Hyperparameters influence the model's complexity, learning ability, and performance. The main hyperparameters used in this study and their tested ranges are shown in Table 3. In this study, Hyperband algorithm (Li et al., 2018) was used at initial stages to optimize the hyperparameters of the LSTM model, followed by manual adjustments based on dataset characteristics to improve the model's ability in capturing fluctuations and rare high peaks more accurately.

195

Hyperparameter	Values
Learning rate	[1e-2, 1e-3]
Activation functions for dense layers	[Tanh, Relu, Leaky relu, Linear]
Activation functions for LSTM layers	[Tanh, Relu]
Number of neurons	20-65
LSTM layers recurrent L2 regularization	[1e-4, 1e-3, 1e-2]
LSTM layers kernel L2 regularization	[1e-4, 1e-3, 1e-2]
Dense layers L2 regularization	[1e-4, 1e-3, 1e-2, 1e-1]
Dropout	[0.01, 0.05, 0.1, 0.2, 0.5]
Number of LSTM Layers	3-5
Number of dense layers	2-3
Loss function loss_coefficient	[1, 1.2, 1.5]

200

Table 3 Ranges of hyperparameters of the model considered for hyperparameter tuning in this study.

2.4.4 Feature importance in LSTM

This study applies the permutation feature importance (PFI) method, which calculates the contribution of each feature to the model's estimation of N_2O emissions. This technique shuffles the value of an individual feature in the dataset, calls the model and observes the degradation of the model performance, which is now producing results with this feature having shuffled values. Therefore, we define the relative importance metric of each feature as the degree of increase in the mean squared error (MSE) in comparison to the MSE in the original model, formulated in Eq. 5:

205

$$\text{Relative importance feature } m = \text{MSE PFI feature } m - \text{MSE original model} \quad (5)$$

210 2.6. Post-processing and statistical analysis

To assess model performance, we used the RMSE and relative RMSE, which is defined as the RMSE divided by the mean value of the target variable in percentage. In addition, for the analysis of the results time series plots for daily emissions as well as cumulative plots for yearly cumulative emissions are created in Matplotlib, Python.

To report the results of different grassland systems in Germany, the model's daily N_2O estimations of individual soil replicates are generated and are finally aggregated on replicates of same treatment.

215



3 Results

3.1 Comparison of model architectures

In this study, two architectures of LSTM were explored as shown in Fig. 2 and architecture A (Fig. 2 a) demonstrated superior performance in capturing general N₂O emission patterns compared to Architecture B (Fig. 2 b). Consequently, the final model is built on architecture A, consisting of an input layer, three LSTM layers, an attention layer, the concatenate layer to add static drivers, a dense layer, a drop out layer and an output layer. The model receives a five-day sequence of temporal drivers, alongside with two static drivers as shown in Eq. 1.

Table 4 compares the RMSE of some of the experiments conducted in this study to identify the optimal model configuration, in which different subsets of soil properties or length of input sequences (5 or 10 days) were tested. For each experiment the hyperparameter tuning is done extensively to obtain an optimum model fitting to the input. In the current dataset, Bulk density (BD) showed strong negative Pearson correlation of -0.9 to both soil organic carbon (SOC) and clay. While all these variables are key drivers of N₂O emissions, only bulk density was included in the model due to slightly better performance in capturing emission peaks. Adding correlated features reduces model's accuracy, likely due to confusing the model with redundant features which do not provide extra information.

The results presented in next section correspond to the best-performing model in catching low and high emissions as well as cumulative emissions (index 1 in Table 4).

Index	Architecture	Input soil properties	Test RMSE ($\mu\text{gm}^{-2}\text{h}^{-1}$)	Test relative RMSE	Sequence length
1	A	BD, pH	18	270%	5
2	A	BD, pH	23	350%	10
3	B	BD, pH	22	340%	5
4	A	pH, SOC	19	290%	5
5	A	SOC, BD, Clay, pH	20	307%	5

Table 4 Model specifications of 5 experiments compared in this study. The models are compared on the basis of the RMSE, the inputs they receive, and the length of the input sequence of days, while the temporal drivers are the same.

3.2 Model performance

The best model performed well in modelling daily emissions in the training set, with an RMSE of $34 \mu\text{gm}^{-2}\text{h}^{-1}$ and a relative RMSE of 200%. The model is also able to successfully estimate daily N₂O emissions in the test set, with an RMSE of $18 \mu\text{gm}^{-2}\text{h}^{-1}$, a relative RMSE of 270% and total bias of -12%. Having relative RMSE higher than 1 is a known effect of the imbalanced dataset. Figure 3 Shows the scatter plots of model performance on the training set, the test set.

The model successfully captures the temporal fluctuations and magnitude of N₂O emissions at the test sites in various instances, although the accuracy varies between soils as shown in Fig. 4. For most sites, the general emission pattern is simulated with a high degree of accuracy. However, in some cases, there is evidence of over- or underestimation of the magnitude of N₂O emissions. For example, at Chamau and Fendt, the model predicted the overall variability of the emissions well, indicating that Chamau experiences higher peaks than other sites; however, in some cases the peak values were not accurately simulated (Fig. 4 a, b and c). At Graswang, the model tended to underestimate fluxes (Fig 4. d).

To assess the generalisability of the model, the Rottenbuch soils (at the Fendt climate) were excluded from the training set and considered as independent test soils. These soils showed variable emissions across years, with peak fluxes exceeding $150 \mu\text{gm}^{-2}\text{h}^{-1}$ during warmer periods, particularly following fertilization events in 2015 and 2016 (Fig. 4 h and i), for which the model captures the patterns in most cases. On the other hand, fertilization during warmer periods in 2017 and 2019 did not result in elevated fluxes, and the model shows significant errors for estimating this behaviour in several cases (Fig. 4 j and k). In general, the results show that N₂O emissions are generally lower during colder seasons, and even fertilization during cold conditions does not lead to highly elevated fluxes; a pattern which is captured well by the model in most cases.

Figure 5 shows the comparison between observed and modelled annual cumulative emissions achieved by aggregating the daily fluxes. The results show that the model creates an accurate estimation of annual emissions for some soils and in different years with a slight error rate, such as Rottenbuch in Fendt 2015



intensive and Rottenbuch in Fendt 2015, 2016, 2017 extensive. But, higher overestimation on annual total emissions are evident for soils such as Rottenbuch in Fendt 2016, 2017, 2020 intensive as well as Chamau (Fig. 5 a and b).

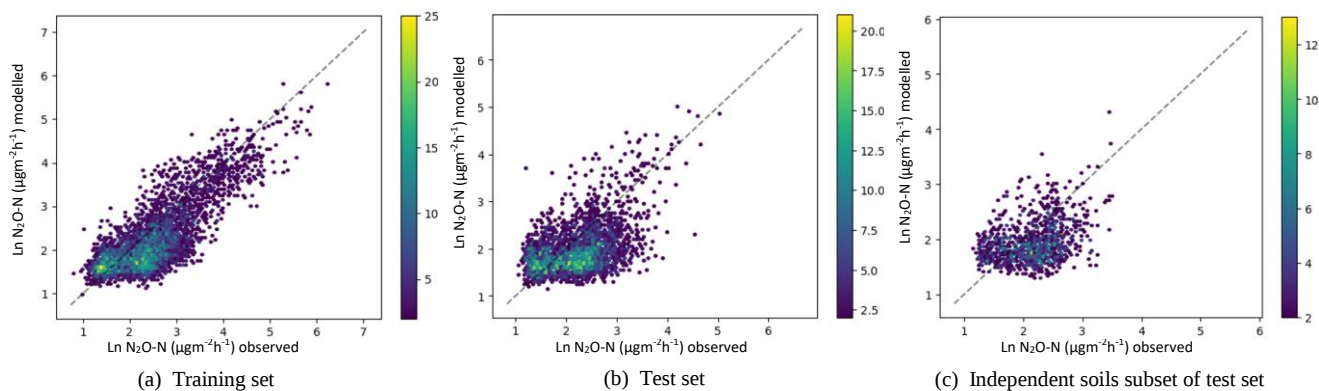


Figure 3 Hexbin plots with density colour bar comparing observed (x-axis) with simulated (y-axis) daily N₂O fluxes on a natural log (ln) scale for (a) the training set and (b) the whole test set, (c) the subset of test set which consists of the independent test soils of Rottenbuch in Fendt. Hexagon colours show the number of predictions falling into them, appearing as lighter (yellow) colours for bins with more data points and darker colours (blue) with less data points.

260 **3.4 Permutation feature importance (PFI)**

Three configurations of PFI analysis were performed on the test set to provide a comprehensive assessment of the relative importance of individual drivers. In the first configuration, PFI was applied on the entire test dataset. This analysis revealed that soil temperature at a depth of 10 cm was the most influential feature, followed by soil pH and fertilization (Fig. 6 a) while all other input features also show a high relative importance. In the second configuration, PFI was applied to individual sites in the test set in order to investigate the the importance of different drivers at site-scale N₂O dynamics. The results indicate that at most sites all temporal drivers were the highly important features for N₂O dynamics (Fig. 6 b-f). However, at the Graswang site BD and pH show a very high importance, together with soil temperature (Fig. 6 d), due to high variability in soil properties within the soil cores of this grassland which affect their N₂O emission (Table 2). Compared to the soil lysimeter systems in Germany including a diverse set of soils with variability in pH and BD, there is only one soil type included at the Chamau site which explains the zero relative importance of these variables at Chamau in Fig. 6 b.

270 In an additional analysis, we studied the varying importance of features across the five-time steps individually, shown in Fig. 7, which further highlights the role of dynamic environmental conditions in driving emissions and reinforces the need to incorporate temporal variability in emission modeling frameworks.

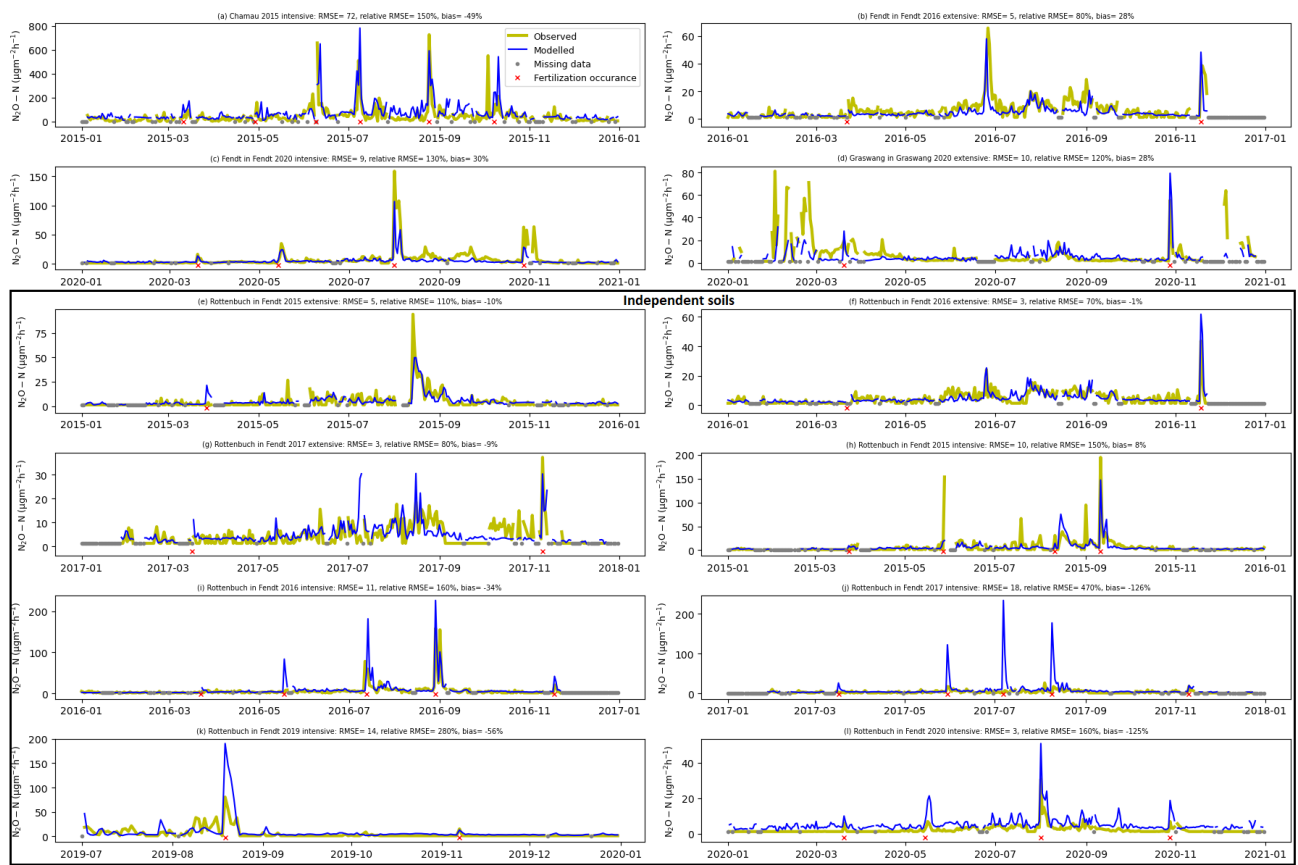


Figure 4 Time series of observed and modelled N_2O fluxes in different soils at different climates and years for the test set including RMSE, relative RMSE and bias for individual soils. Note that the axis scales differ between panels to enable visibility of the data and visualize the differences between N_2O emission rates at different sites in the experiment.

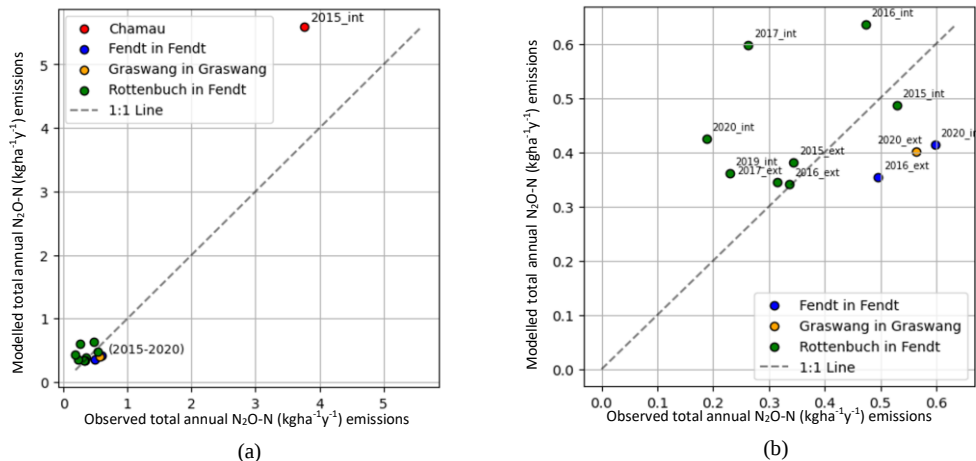
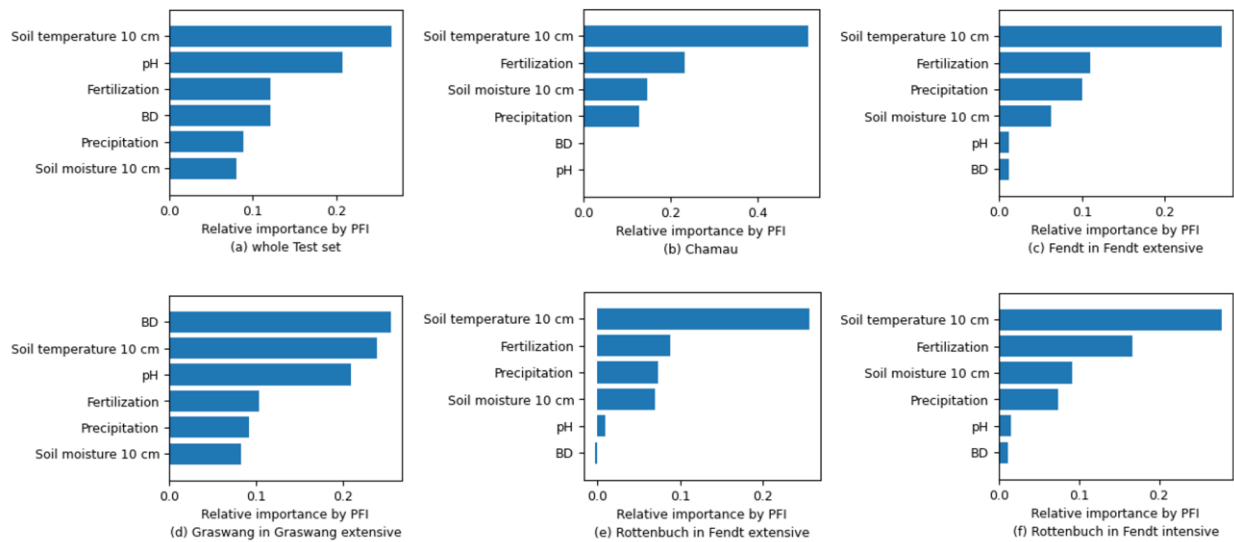


Figure 5 Cumulative observed and modelled annual N_2O emissions across multiple years for different sites representing various soils in climates: (a) all soils combined (b) enlarged version showing only the German soils including the independent test soil, Rottenbuch in Fendt.



285

Figure 6 Relative feature importance (x-axis) of drivers (y-axis) in the test set based on MSE calculated by PFI for the complete test set and individual sites.

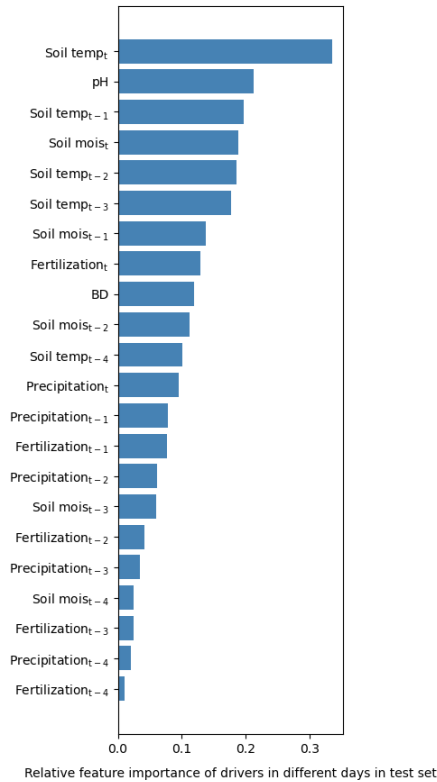


Figure 7 PFI feature importance at individual days, in which t index indicates the current day.



4 Discussion

290 In this study, a long short-term memory (LSTM) neural network was employed to model daily N₂O emissions from grasslands in Central Europe, based on soil properties as well as key environmental and management drivers over a 5-days window.

Our study makes a novel contribution to simulating soil N₂O emission from grassland soils using ML approaches. This is achieved by building a comprehensive problem setup, collecting a large and diverse dataset of 28,000 data points which encompasses a wide range of soil properties and climatic conditions (Table 1 and 2), incorporating an independent test site, and accounting for the role of environmental changes on N₂O emissions within several
 295 days in a sequential setup modelled by the LSTM. By contrast, recent studies (Table 5) which used ML to model GHG emissions have generally relied on smaller datasets from single sites and many of them used classical machine learning approaches with non-sequential data.

Adjuik et al. (2022) used data of variable cropping systems, sites and fertilization amendments to estimate CO₂ emissions using minimal predictors. However, train and test set split were performed randomly on the whole dataset meaning that there was no independent test site available for evaluation. Similarly, Jiang et al. (2023) collected a diverse global dataset, using 52 studies of CH₄ and N₂O from paddy fields, and applied an ensemble model combining classical ML
 300 approaches to simulate daily emissions of a single site (3 years paddy field experiment in Kushan, China) on 558 datapoints (Jiang et al., 2023). In another study, Saha et al. (2021) used a classical ML approach to model sub-daily N₂O emissions from two fertilised maize systems in the US, including an independent test site located 2 km from the training sites, but with a significantly smaller dataset and less variability in climatic conditions.

In only one other study, LSTM was used to simulate CO₂ and N₂O emissions for a single site in Quebec, Canada by Hamrani et al., 2020, and was compared with a set of ML algorithms in 3 categories: classical ML, shallow learning and other deep learning algorithms. This study demonstrated that only the LSTM
 305 model was able to provide a good prediction of N₂O fluxes, and was also more accurate than the process-based Root Zone Water Quality Model (Hamrani et al., 2020).

Study	Gas	Data source	Data size	Input drivers	Land type	ML algorithms
Current study	N ₂ O	-Germany, Switzerland, Austria -10 years	28000	Soil temperature, soil moisture, precipitation, occurrence of fertilization, soil pH, soil BD	Grassland	-LSTM (deep learning)
Adjuik et al., 2022	CO ₂	Multiple sites from GRACENET	7863	Soil temperature, Air temperature, fertilizer amendment class, soil classification, and crop type	Cropland	-Random forest -Support vector regression -Gradient boosting (best)
Jiang et al., 2023	CH ₄ , N ₂ O	-Kunshan, China -Three years	CH ₄ : 552 N ₂ O: 558	Soil temperature, soil moisture or Irrigation volume, air temperature (maximum, minimum, and average), soil redox potential, rainfall, wind speed, and radiation	Paddy fields	-Random forest -K-nearest neighbours -Gradient boosting -Linear regression -Ensemble model of all (best)
Jiang et al., 2023	CH ₄ , N ₂ O	-Cumulative fluxes global dataset collected from 52 studies	CH ₄ : ~60 N ₂ O: ~60	Soil organic carbon, pH, texture, inorganic nitrogen fertilizer application, and non-flooded days	Paddy fields	-Ensemble model of random forest, k-nearest neighbours, gradient boosting and linear regression
Saha et al., 2021	N ₂ O	-2 sites in US Midwest -Six years	2246	Soil moisture, air temperature, days after fertilization, soil texture, soil carbon, precipitation, nitrogen (N) fertilizer rate or inorganic N availability in the soil	Corn rotation	-Random forest model in conjunction with a process-based model (SALUS)
Hamrani et al., 2020	CO ₂ , N ₂ O	-Quebec, Canada -Five years	126	soil temperature, soil volumetric water content, air temperature, precipitation and humidity, wind speed, surface pressure, and crop nitrogen uptake	Drainage and water table management research site	-LSTM (deep learning) -Random forest (classical ML) -Feed-forward neural networks (shallow learning)

Table 5 Review of several studies which applied different ML approaches for modelling soil N₂O emissions



4.2 Model performance and role of input drivers

Analysis of the dataset revealed the interaction between the selected input variables for the model. For instance, even under favourable temperature conditions (warmer periods), fertilization alone did not consistently result in immediate N₂O peaks, which suggests that increased nitrogen availability, leads to increased emissions only when also other environmental conditions such as soil moisture are conducive. Another important aspect about the N₂O emissions is that the duration of elevated emissions is influenced by the interplay of environmental variables. For example, Fig. 4 c, h and j show instances of continued changes in N₂O emission magnitudes extending over multiple days following fertilization events at different soils in August which is also captured by the LSTM model.

In addition to the emission peaks following fertilization, several high peaks were observed on days with no fertilization, such as in Fendt in Fendt extensive in June 2016 (Fig. 4 b) as well as Graswang in Graswang extensive in January 2020 (Fig. 4 d). In the first case, the model successfully simulated the peak (earlier timing) caused by changes in the soil temperature in conductive soil moisture, but in the latter one the model simulated fluctuations but underestimated the magnitudes of the peaks.

Towards developing the core model, we conducted a series of experiments to assess the impact of varying input sequence length and feature set on model performance, in which sequence lengths of 5, 7 and 10 days were tested. While longer sequences have the potential to capture long-term soil processes, their benefits were not evident in the current dataset, likely due to the dataset's limited occurrence of such events. For example, in the current dataset there are very rare cases where the high peaks after fertilizations occur later than 5 days. This suggests that the value of longer sequences may become more apparent when they are applied to larger and more diverse datasets. We also tested an expanded feature set to incorporate other important drivers of N₂O emissions such as the rate of fertilizer nitrogen (total or inorganic N), which did not improve the results due limited variability across sites and events, offering little additional information to improve predictive accuracy. However, if the model is to be applied to sites with more variable fertilizer practices, including explicit N application rates would likely enhance model performance.

4.3 Feature importance analysis

The PFI analysis on the whole test set and individual soils, consistently demonstrate that all input features contribute significantly to model performance (Fig. 6). The substantial importance of both soil and meteorological features indicates that the model captured not only the meteorological effects, but also the role of soil parameters. Soil temperature, ranks the most important features which aligns with previous studies highlighting its critical role in regulating microbial processes (Wang et al., 2021; Butterbach-Bahl et al., 2013; Hamrani et al., 2020). Soil temperature might often be unavailable at many sites around the world, but in case of high Pearson correlation with maximum air temperature (0.86 in our dataset), the latter can be used as a substitute with minimal impact on model performance.

Fertilization events lead to pronounced short-term peaks in emissions, however, because in our study such events are infrequent, occurring maximum of five times per year, their permutation does not result a significant decline on model's overall performance comparing to more consistently present features. Nevertheless, fertilization still shows relatively high importance in PFI analysis and Fig. 4 confirms that the model effectively captures the associated emission peaks.

Soil pH and BD also appear as key drivers in the general PFI analyses, likely due to their influence on soil-specific emission patterns. BD serves as a proxy for soil porosity and compaction, which are factors known to influence oxygen availability strongly impacting microbial activity and importance of nitrification and denitrification processes (Li et al., 2024; Butterbach-Bahl et al., 2013). Moreover, pH values affect the microbial processes as near neutral soil pH values (6.6 - 8.3) are optimal for denitrification (Šimek et al. 2002), while low soil pH (<5.5) inhibits N₂O reductase, the enzyme that converts N₂O to N₂ during denitrification (Samad et al., 2016).



Despite their lower rank in Fig. 6 a, soil moisture and precipitation remain influential showing considerable relative importance, suggesting their added value for estimating N₂O emissions. Fig. 7 proves that soil moistures at three-days window are among top 10 influential features in model's estimations. An expanded dataset with more variety in emission patterns which are impacted by these two drivers could highlight their roles further.

4.4 Comparison to process-based models

Previous studies have evaluated N₂O dynamics with biogeochemical models using data from laboratory experiments and grassland sites under different climatic and soil conditions.

Zimmermann et al. (2018) found relative root mean square error (RMSE) values ranging from 140–234% in DayCent, 261–503% in DNDC-9.4, 352–652% in DNDC-9.5, and 160–2079% in ECOSSE based on daily data for 4 sites in the Republic of Ireland. Fitton et al. (2014) found relative RMSE values of 143%, 212%, and 423 % for DayCent across 3 grassland sites in United Kingdom. By comparison, the relative RMSE for the LSTM model in our study is in average 270% across the full test set. However, for 80% percent of site-years the relative RMSE ranged from 67% to 160%. Only in 2017 for Rottenbuch in Fendt soil the RMSE was 467% due to poor model performance in estimating the unusually low emissions following fertilization (Fig. 4j).

At the Chamau site, Fuchs et al. (2020) validated biogeochemical models, finding relative RMSE values of 168–173% for DayCent, 221% for PaSim, and 158–161% for APSIM. These values are slightly higher than the RMSE of the LSTM simulations for Chamau in our study (148%). Similarly, Denk et al., 2019 reported the RMSE of the biogeochemical model LandscapeDNDC for Chamau during summer 2013 as 116.66 $\mu\text{g m}^{-2}\text{h}^{-1}$ of N₂O-N, higher than the RMSE of 55 $\mu\text{g m}^{-2}\text{h}^{-1}$ of our LSTM model during the same period. In addition, in some instances, LDNDC showed a delay of two days in capturing emissions peaks after fertilization or rewetting, which is a known issue also reported by Gaillard et al., 2018. In our LSTM model, although timing errors occurred in some instances (e.g., peak emissions in June, Fig. 4a–b), but the model generally captured the timing, continuity and magnitude of emission peaks correctly in many other cases (Fig. 4) so the delays are not characterized as a general issue in the LSTM simulations

Bias values also varied widely across biogeochemical models. Fitton et al. (2014) found relative biases for DayCent ranging from –84% to +10% for the UK grassland sites. Zimmermann et al. (2018) reported relative biases for Irish grassland sites, ranging from –5% to 88% for DayCent, –116% to –71% for DNDC-9.4, –48% to 87% for DNDC 9.5, and –1395% to 40% for ECOSSE. At Chamau, Fuchs et al. (2020) reported moderate values of –35% to +15% for DayCent, +41% for PaSim and –47% to –43% for APSIM. By comparison, the LSTM model in our study showed bias values ranging from –126% to +30% across soils in different years (Fig. 4). The extreme value (–126%) again corresponded to the poor performance in 2017 at Rottenbuch in Fendt soil (Fig. 4j). However, in other years at the same soil (Fig. 4e–h), bias values were much lower (–1.61% to +10%).

Overall, the results show that the LSTM model can be promising alternative compared to established process-based biogeochemical models for predicting N₂O emissions from grassland soils, providing robust performance in RMSE and bias for most site-years, especially in correctly capturing emission magnitude and timing.

4.5 Implications and limitations

The performance of the LSTM model in capturing daily emission of the test set suggests that the model has potential for both gap-filling and predictive applications in similar alpine grassland environments, with possible utility for upscaling emissions across comparable regions. However, several limitations should be acknowledged. A key challenge with LSTM models lies in their limited interpretability compared with non-recurrent models due to complex recurrent structure. Moreover, the dataset exhibits a class imbalance, as it includes infrequent but high-magnitude emission peaks alongside prolonged periods of low emissions. This imbalance complicates the model's ability to learn the drivers of extreme events.

While this and several other studies have demonstrated the promise of machine learning (ML) models for simulating soil greenhouse gas (GHG) emissions, the full potential of these approaches remains underexplored. We identify three key directions for future research that could further advance the application of ML in this domain. Firstly, expanding the dataset to encompass more geographically and climatically diverse sites, along with a wider range of management practices would allow for the development of more generalized N₂O emission models for grasslands across different agroecosystems. Secondly, recent advancements in deep learning models for time series forecasting, such as transformer neural networks (Vaswani et al., 2017), could be leveraged in combination with better explainable AI tools to further enhance the progress in development of the generalized model. Transformers have demonstrated

significant potential in time series modelling because of their ability to capture long-range dependencies and handle multivariate data effectively (Zerveas et al., 2021). Unlike traditional recurrent models such as LSTM, transformers process entire sequences in parallel, making them highly scalable and efficient (Vaswani et al., 2017). Finally, other versions of the machine learning models which are trained with air temperature and precipitation instead of soil temperature and moisture could be assessed using a more diverse dataset, enabling N₂O emission predictions even where direct soil measurements are missing.

5 Conclusion

This study demonstrates the effectiveness of an LSTM model in capturing temporal dynamics of daily N₂O emission data of several grassland sites in Central Europe, highlighting the model's suitability for predicting soil N₂O emissions. By collecting a unique dataset from multiple grasslands sites and treatments, we developed a generic model capable of accurately predicting emissions across diverse conditions. Notably, the model's performance remained robust when tested on independent soil that was not used for training, suggesting good transferability. This model can therefore be used not only as a gap-filling model, but potentially to regionally upscale emissions to alpine grasslands for which the input measurements are available. Future work should aim to: (1) expand the dataset to include a broader and more diverse dataset, both in terms of geographic coverage and input drivers, enhancing the model's robustness and applicability at larger scales; and (2) explore recent advances in deep learning, such as transformer models, to further improve prediction accuracy and model interpretability.

Appendix A

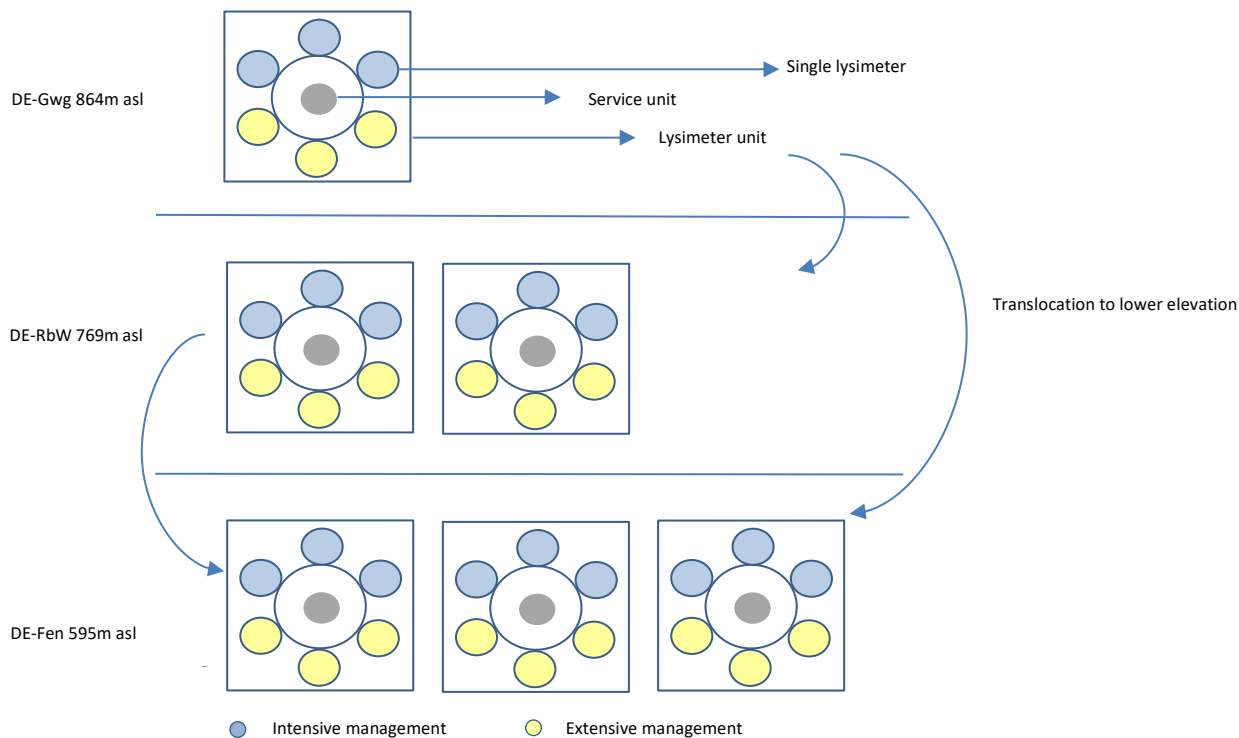


Figure A1 Scheme of lysimeter replicates originated or translocated in different grasslands, Graswang (DE-Gwg), Rottenbuch (DE-RbW), and Fendt (DE-Fen) within the TERENO Pre-Alpine Observatory in Germany (Kiese et al., 2018).



420

Soil	Replicate-Management	Mean	Std	Min	Max
Neustift		20.87	6.08	7.94	37.89
Chamau		43.84	9.02	16.93	54.98
Fendt in Fendt	L1-Intensive	42.04	6.78	17.85	54.97
Fendt in Fendt	L2-Intensive	40.34	6.24	18.22	51.47
Fendt in Fendt	L3-Intensive	41.64	6.99	17.85	54.58
Fendt in Fendt	L1-Extensive (2014-2018), Intensive (2019-2020)	42.96	6.67	17.85	53.90
Fendt in Fendt	L2-Extensive (2014-2018), Intensive (2019-2020)	40.09	7.44	16.09	53.90
Fendt in Fendt	L3-Extensive (2014-2018), Intensive (2019-2020)	43.51	7.86	17.85	54.59
Graswang in Fendt	L1-Intensive	44.44	7.22	17.28	56.14
Graswang in Fendt	L2-Intensive	43.54	6.92	19.37	53.39
Graswang in Fendt	L3-Intensive	46.25	5.90	21.01	53.56
Graswang in Fendt	L1-Extensive	45.76	7.40	19.60	57.05
Graswang in Fendt	L2-Extensive	43.64	7.04	14.90	57.65
Graswang in Fendt	L3-Extensive	40.78	6.74	19.60	54.34
Rottenbuch in Fendt	L1-Intensive	41.59	6.34	16.90	51.03
Rottenbuch in Fendt	L2-Intensive	38.97	6.25	15.56	51.70
Rottenbuch in Fendt	L3-Intensive	43.36	6.58	17.07	52.79
Rottenbuch in Fendt	L1-Extensive (2014-2018), Intensive (2019-2020)	41.68	5.55	17.68	50.20
Rottenbuch in Fendt	L2-Extensive (2014-2018), Intensive (2019-2020)	42.42	7.13	15.94	50.67
Rottenbuch in Fendt	L3-Extensive (2014-2018), Intensive (2019-2020)	44.05	7.26	17.63	53.69
Graswang in Graswang	L1-Intensive	47.15	7.80	13.38	58.15
Graswang in Graswang	L2-Intensive	46.73	5.41	18.37	53.28
Graswang in Graswang	L3-Intensive	48.05	6.20	17.28	55.53
Graswang in Graswang	L1-Extensive	47.89	5.78	16.98	55.85
Graswang in Graswang	L2-Extensive	49.08	6.42	19.37	57.94
Graswang in Graswang	L3-Extensive	47.89	5.63	21.19	55.85

(a) soil moisture statistics of the soils

Soil	Replicate-Management	Mean	Std	Min	Max
Neustift		9.36	6.76	0.31	19.41
Chamau		12.09	6.46	1.18	25.08
Fendt in Fendt	L1-Intensive	10.80	6.82	0.49	24.40
Fendt in Fendt	L2-Intensive	10.55	6.91	0.25	24.38
Fendt in Fendt	L3-Intensive	11.02	6.86	0.40	24.42
Fendt in Fendt	L1-Extensive (2014-2018), Intensive (2019-2020)	10.87	6.98	0.23	25.09
Fendt in Fendt	L2-Extensive (2014-2018), Intensive (2019-2020)	10.88	6.74	0.56	24.31
Fendt in Fendt	L3-Extensive (2014-2018), Intensive (2019-2020)	10.92	6.91	-0.77	24.82
Graswang in Fendt	L1-Intensive	10.53	6.87	0.63	24.74
Graswang in Fendt	L2-Intensive	10.73	6.74	0.45	24.79
Graswang in Fendt	L3-Intensive	10.86	6.80	-0.44	24.59
Graswang in Fendt	L1-Extensive	10.79	6.86	0.14	24.31
Graswang in Fendt	L2-Extensive	11.42	6.66	1.02	25.14
Graswang in Fendt	L3-Extensive	10.68	6.80	-1.28	24.38
Rottenbuch in Fendt	L1-Intensive	10.83	6.72	-0.36	25.14
Rottenbuch in Fendt	L2-Intensive	11.07	6.97	-0.22	24.96
Rottenbuch in Fendt	L3-Intensive	10.75	6.76	-0.30	24.58
Rottenbuch in Fendt	L1-Extensive (2014-2018), Intensive (2019-2020)	10.48	6.59	-0.26	24.92
Rottenbuch in Fendt	L2-Extensive (2014-2018), Intensive (2019-2020)	11.16	6.56	0.74	24.40
Rottenbuch in Fendt	L3-Extensive (2014-2018), Intensive (2019-2020)	10.82	6.68	-0.36	24.48
Graswang in Graswang	L1-Intensive	9.40	6.93	-1.35	23.06
Graswang in Graswang	L2-Intensive	8.89	6.70	-1.35	22.20
Graswang in Graswang	L3-Intensive	9.31	6.80	-1.21	22.32
Graswang in Graswang	L1-Extensive	9.34	6.64	-0.71	22.64
Graswang in Graswang	L2-Extensive	9.28	6.50	-1.06	22.18
Graswang in Graswang	L3-Extensive	9.45	6.62	-1.43	22.73

(b) soil temperature statics of the soils

425

430



Soil	Replicate-Management	Mean	Std	Min	Max
Neustift		12.54	10.41	0.11	48.86
Chamau		70.11	114.03	-739.12	1155.83
Fendt in Fendt	L1-Intensive	5.84	13.56	-25.01	248.33
Fendt in Fendt	L2-Intensive	4.52	7.79	-26.17	160.07
Fendt in Fendt	L3-Intensive	11.44	30.64	-25.68	707.50
Fendt in Fendt	L1-Extensive (2014-2018), Intensive (2019-2020)	6.27	19.81	-30.72	357.82
Fendt in Fendt	L2-Extensive (2014-2018), Intensive (2019-2020)	6.13	12.79	-15.00	192.32
Fendt in Fendt	L3-Extensive (2014-2018), Intensive (2019-2020)	14.97	48.18	-24.85	750.38
Graswang in Fendt	L1-Intensive	7.16	35.14	-25.57	960.49
Graswang in Fendt	L2-Intensive	5.77	14.12	-26.52	288.06
Graswang in Fendt	L3-Intensive	6.10	16.39	-20.97	326.71
Graswang in Fendt	L1-Extensive	6.09	17.05	-28.28	428.09
Graswang in Fendt	L2-Extensive	4.88	8.21	-26.73	164.57
Graswang in Fendt	L3-Extensive	3.52	5.80	-27.29	79.14
Rottenbuch in Fendt	L1-Intensive	4.30	10.84	-30.25	226.26
Rottenbuch in Fendt	L2-Intensive	4.40	13.25	-22.33	275.85
Rottenbuch in Fendt	L3-Intensive	6.09	16.04	-10.27	323.24
Rottenbuch in Fendt	L1-Extensive (2014-2018), Intensive (2019-2020)	2.84	4.03	-15.93	52.58
Rottenbuch in Fendt	L2-Extensive (2014-2018), Intensive (2019-2020)	4.09	6.15	-14.12	81.62
Rottenbuch in Fendt	L3-Extensive (2014-2018), Intensive (2019-2020)	5.68	10.92	-13.35	132.58
Graswang in Graswang	L1-Intensive	14.14	36.82	0.22	513.37
Graswang in Graswang	L2-Intensive	11.78	26.53	1.19	230.11
Graswang in Graswang	L3-Intensive	9.01	22.17	-0.48	283.05
Graswang in Graswang	L1-Extensive	11.37	21.43	1.14	222.33
Graswang in Graswang	L2-Extensive	8.33	15.67	0.46	103.26
Graswang in Graswang	L3-Extensive	4.90	5.96	1.12	49.66

435 (c) N₂O statistics of the soils

Table A1 Statistics of the soils in different climates with respect to (a) soil moisture at 0–10 cm depth, (b) soil temperature at 0–10 cm depth, and (c) daily N₂O emissions.

Data and code availability

440 The datasets analysed during the current study were sourced from existing publications. The code to the implemented model in python could be provided by the corresponding authors upon request.

Author contributions

SP conceptualization, investigation, data collection and analysis, model implementation, visualization, writing -original draft preparation, -review and editing. KF conceptualization, writing -review and editing, supervision. BW: data provision, writing -review and editing. RK: data provision, writing -
 445 review and editing, supervision. CS conceptualization, writing -review and editing, supervision, funding acquisition, project administration.

Competing interests: The contact author has declared that none of the authors has any competing interests.

Acknowledgment

We thank George Wohlfahrt, Lukas Hörtnagl, Nina Buchmann and Iris Feigenwinter (University of Innsbruck and ETH Zurich) and their funders for supporting us in compiling the datasets from Austria and Switzerland.

450



References

- 455
- Abdalla, M., Jones, M., Yeluripati, J., Smith, P., Burke, J., & Williams, M. (2010). Testing DayCent and DNDC model simulations of N₂O fluxes and assessing the impacts of climate change on the gas flux and biomass production from a humid pasture. *Atmospheric Environment*, 44(25), 2961–2970. <https://doi.org/10.1016/j.atmosenv.2010.05.018>
 - Adjuik, T. A., & Davis, S. C. (2022). Machine learning approach to simulate soil CO₂ fluxes under cropping systems. *Agronomy*, 12, 197. <https://doi.org/10.3390/agronomy12010197>
- 460
- Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *Proc. 3rd Int. Conf. Learn. Representations (ICLR)*. <https://arxiv.org/abs/1409.0473>
 - Bigaignon, L., Fieuzal, R., Delon, C., & Talleg, T. (2020). Combination of two methodologies, artificial neural network and linear interpolation, to gap-fill daily nitrous oxide flux measurements. *Agricultural and Forest Meteorology*, 291, 108037. <https://doi.org/10.1016/j.agrformet.2020.108037>
- 465
- Brilli, L., Bechini, L., Bindi, M., Carozzi, M., Cavalli, D., Conant, R., Dorich, C. D., Doro, L., Ehrhardt, F., Farina, R., Ferrise, R., Fitton, N., Francaviglia, R., Grace, P., Iocola, I., Klumpp, K., Léonard, J., Martin, R., Massad, R. S., Recous, S., Seddaiu, G., Sharp, J., Smith, P., Smith, W. N., Soussana, J.-F., & Bellocchi, G. (2017). Review and analysis of strengths and weaknesses of agro-ecosystem models for simulating C and N fluxes. *Science of the Total Environment*, 598, 445–470. <https://doi.org/10.1016/j.scitotenv.2017.03.208>
- 470
- Butterbach-Bahl, K., Baggs, E. M., Dannenmann, M., Kiese, R., & Zechmeister-Boltenstern, S. (2013). Nitrous oxide emissions from soils: How well do we understand the processes and their controls? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368, 20130122. <https://doi.org/10.1098/rstb.2013.0122>
 - Chataut, G., Bhatta, B., Joshi, D., Subedi, K., & Kafle, K. (2023). Greenhouse gases emission from agricultural soil: A review. *Journal of Agriculture and Food Research*, 11, 100533. <https://doi.org/10.1016/j.jafr.2023.100533>
- 475
- Chhajaj, P., Shah, M., & Kshirsagar, A. (2022). The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction. *Decision Analytics Journal*, 2, 100015. <https://doi.org/10.1016/j.dajour.2021.100015>
 - Chollet, F., et al. (2015). Keras. <https://keras.io>
 - Denk, T. R. A., D. Kraus, Kiese, R., Butterbach-Bahl, K., and Wolf, B. (2019). Constraining N cycling in the ecosystem model LandscapeDNDC with the stable isotope model SIMONE. *Ecology* 100 (5): e02675. <https://doi.org/10.1002/ecy.2675>
- 480
- Ehrhardt, F., et al. (2018). Assessing uncertainties in crop and pasture ensemble model simulations of productivity and N₂O emissions. *Global Change Biology*, 24, e603–e616. 10.1111/gcb.13965
 - Feigenwinter, I., Hörtnagl, L., & Buchmann, N. (2023). N₂O and CH₄ fluxes from intensively managed grassland: The importance of biological and environmental drivers vs. management. *Science of the Total Environment*, 903, 166389. <https://doi.org/10.1016/j.scitotenv.2023.166389>
- 485
- Fitton, N., Datta, A., Hastings, A., Kuhnert, M., Topp, C. F. E., Cloy, J. M., et al. (2014). The challenge of modelling nitrogen management at the field scale: Simulation and sensitivity analysis of N₂O fluxes across nine experimental sites using DailyDayCent. *Environmental Research Letters*, 9(9), 095003. <https://doi.org/10.1088/1748-9326/9/9/095003>
- 490
- Fuchs, K., Hörtnagl, L., Buchmann, N., Eugster, W., Snow, V., & Merbold, L. (2018). Management matters: Testing a mitigation strategy for nitrous oxide emissions using legumes on intensively managed grassland. *Biogeosciences*, 15, 5519–5534. <https://doi.org/10.5194/bg-15-5519-2018>
 - Fuchs, K., Merbold, L., Buchmann, N., Bretscher, D., Brilli, L., Fitton, N., et al. (2020). Multimodel evaluation of nitrous oxide emissions from an intensively managed grassland. *Journal of Geophysical Research: Biogeosciences*, 125, e2019JG005261. <https://doi.org/10.1029/2019JG005261>



- Gilgen, A. K., & Buchmann, N. (2009). Response of temperate grasslands at different altitudes to simulated summer drought differed but scaled with annual precipitation. *Biogeosciences*, 6(11), 2525–2539.
- 495 • Giltrap, D. L., Li, C., & Saggarr, S. (2010). DNDC: A process-based model of greenhouse gas fluxes from agricultural soils. *Agriculture, Ecosystems & Environment*, 136, 292–300. <https://doi.org/10.1016/j.agee.2009.07.006>
- Hamrani, A., Akbarzadeh, A., & Madramootoo, C. A. (2020). Machine learning for predicting greenhouse gas emissions from agricultural soils. *Science of the Total Environment*, 741, 140338. <https://doi.org/10.1016/j.scitotenv.2020.140338>
- 500 • Hörtnagl, L., Barthel, M., Buchmann, N., Eugster, W., Butterbach-Bahl, K., Díaz-Pinés, E., Zeeman, M., Klumpp, K., Kiese, R., Bahn, M., Hammerle, A., Lu, H., Ladreiter-Knauss, T., Burri, S., & Merbold, L. (2018). Greenhouse gas fluxes over managed grasslands in Central Europe. *Global Change Biology*, 24, 1843–1872. <https://doi.org/10.1111/gcb.14079>
- Hörtnagl, L., & Wohlfahrt, G. (2014). Methane and nitrous oxide exchange over a managed hay meadow. *Biogeosciences*, 11, 7219–7236. <https://doi.org/10.5194/bg-11-7219-2014>
- 505 • Hörtnagl, L., Feigenwinter, I., Wang, Y., & others / Project leader: Nina Buchmann (2025). *Eddy covariance ecosystem fluxes, meteorological data and detailed management information for the intensively managed grassland site Chamau in Switzerland, collected between 2005 and 2024*. Dataset, ETH Zürich. <https://doi.org/10.3929/ethz-b-000747025>
- Khalil, M. I., Abdalla, M., Lanigan, G., Osborne, B., & Müller, C. (2016). Evaluation of parametric limitations in simulating greenhouse gas fluxes from Irish arable soils using three process-based models. *Agricultural Sciences*, 07(08), 503–520. <https://doi.org/10.4236/as.2016.78051>
- 510 • Kiese, R., Fersch, B., Baessler, C., Brosy, C., Butterbach-Bahl, K., Chwala, C., Dannenmann, M., Fu, J., Gasche, R., Grote, R., Jahn, C., Klatt, J., Kunstmann, H., Mauder, M., Rödiger, T., Smiatek, G., Soltani, M., Steinbrecher, R., Völksch, I., Werhahn, J., Wolf, B., Zeeman, M., & Schmid, H. P. (2018). The TERENO Pre-Alpine Observatory: Integrating meteorological, hydrological, and biogeochemical measurements and modelling. *Vadose Zone Journal*, 17, 1–17. <https://doi.org/10.2136/vzj2018.03.0060>
- Kopittke, P. M., Dalal, R. C., McKenna, B. A., Smith, P., Wang, P., Weng, Z., van der Bom, F. J. T., & Menzies, N. W. (2024). Soil is a major contributor to global greenhouse gas emissions and climate change. *EGUsphere* [preprint]. <https://doi.org/10.5194/egusphere-2024-1782>
- 515 • Le, X.-H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of Long Short-Term Memory (LSTM) Neural Network for flood forecasting. *Water*, 11(7), 1387. <https://doi.org/10.3390/w11071387>
- Li, C., Frolking, S., & Frolking, T. A. (1992). A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity. *Journal of Geophysical Research: Atmospheres*, 97, 9759–9776. <https://doi.org/10.1029/92JD00509>
- 520 • Li, D., Li, Y., Yao, S., Zhou, H., Huang, S., Peng, X., & Meng, Y. (2024). Dynamics of nitrogen mineralization and nitrogen cycling functional genes in response to soil pore size distribution. *European Journal of Soil Biology*, 123, 103692. <https://doi.org/10.1016/j.ejsobi.2024.103692>
- Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2016). Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. *Journal of Machine Learning Research*, 18(558), 1–52. <https://arxiv.org/abs/1603.06560>
- Li, Y., Chen, D., Zhang, Y., Edis, R., & Ding, H. (2005). Comparison of three modeling approaches for simulating denitrification and nitrous oxide emissions from loam-textured arable soils: Modeling denitrification and nitrous oxide. *Global Biogeochemical Cycles*, 19, GB3002. <https://doi.org/10.1029/2004GB002392>
- 525 • Lin, C., Yeap, T., & Kiringa, I. (2022). Stacked Bidirectional LSTM for predicting emission of nitrous oxide. In *Proceedings of the Canadian Conference on Artificial Intelligence*. <https://doi.org/10.21428/594757db.daad1be1>
- Merbold, L., Eugster, W., Stieger, J., Zahniser, M., Nelson, D., & Buchmann, N. (2014). Greenhouse gas budget (CO₂, CH₄ and N₂O) of intensively managed grassland following restoration. *Global Change Biology*, 20(6), 1913–1928. <https://doi.org/10.1111/gcb.12518>



- 530 • Parton, W. J., Hartman, M., Ojima, D., & Schimel, D. (1998). DAYCENT and its land surface submodel: Description and testing. *Global and Planetary Change*, 19, 35–48. [https://doi.org/10.1016/S0921-8181\(98\)00040-X](https://doi.org/10.1016/S0921-8181(98)00040-X)
- Parton, W. J., Mosier, A. R., Ojima, D. S., Valentine, D. W., Schimel, D. S., Weier, K., & Kulmala, A. E. (2001). Generalized model for NO_x and N₂O emissions from soils. *Global Biogeochemical Cycles*, 10(3), 401–412. <https://doi.org/10.1029/96GB01455>
- 535 • Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://doi.org/10.48550/arXiv.1201.0490>
- Pütz, T., Kiese, R., Wollschläger, U., et al. (2016). TERENO-SOILCan: A lysimeter-network in Germany observing soil processes and plant diversity influenced by climate change. *Environmental Earth Sciences*, 75, 1242. <https://doi.org/10.1007/s12665-016-6031-5>
- Roth, K. (2006). Bodenkartierung und GIS-basierte Kohlenstoffinventur von Graslandböden (diploma thesis). University of Zurich, Zurich, Switzerland
- 540 • Saha, D., Basso, B., & Robertson, G. P. (2021). Machine learning improves predictions of agricultural nitrous oxide (N₂O) emissions from intensively managed cropping systems. *Environmental Research Letters*, 16. <https://doi.org/10.1088/1748-9326/abd2f3>
- Saggar, S., Jha, N., Deslippe, J., Bolan, N. S., Luo, J., Giltrap, D. L., Kim, D.-G., Zaman, M., & Tillman, R. W. (2013). Denitrification and N₂O:N₂ production in temperate grasslands: Processes, measurements, modelling and mitigating negative impacts. *Science of the Total Environment*, 465, 173–195. <https://doi.org/10.1016/j.scitotenv.2013.01.050>
- 545 • Samad, M. S., Bakken, L. R., Nadeem, S., Clough, T. J., de Klein, C. A. M., Richards, K. G., et al. (2016). High-resolution denitrification kinetics in pasture soils link N₂O emissions to pH, and denitrification to C mineralization. *PLoS ONE*, 11(3), e0151713. <https://doi.org/10.1371/journal.pone.0151713>
- Sánchez-Martín, L., Vallejo, A., Dick, J., & Skiba, U. M. (2008). The influence of soluble carbon and fertilizer nitrogen on nitric oxide and nitrous oxide emissions from two contrasting agricultural soils. *Soil Biology and Biochemistry*, 40, 142–151. <https://doi.org/10.1016/j.soilbio.2007.07.016>
- 550 • Smith, P., et al. (2020). The role of soils in delivering Nature's Contributions to People. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375, 20190122. <https://doi.org/10.1098/rstb.2019.0122>
- Šimek, M., Jiřová, L., & Hopkins, D. W. (2002). What is the so-called optimum pH for denitrification in soil? *Soil Biology and Biochemistry*, 34(9), 1227–1234. [https://doi.org/10.1016/S0038-0717\(02\)00059-7](https://doi.org/10.1016/S0038-0717(02)00059-7)
- 555 • Soltani, M., Mauder, M., Laux, P., et al. (2018). Turbulent flux variability and energy balance closure in the TERENO prealpine observatory: A hydrometeorological data analysis. *Theoretical and Applied Climatology*, 133, 937–956. <https://doi.org/10.1007/s00704-017-2235-z>
- Stehfest, E., & Müller, C. (2004). Simulation of N₂O emissions from a urine-affected pasture in New Zealand with the ecosystem model DayCent: Simulation of N₂O from grassland. *Journal of Geophysical Research*, 109, D03109. <https://doi.org/10.1029/2003JD004261>
- 560 • Thorburn, P. J., Biggs, J. S., Collins, K., & Probert, M. E. (2010). Using the APSIM model to estimate nitrous oxide emissions from diverse Australian sugarcane production systems. *Agriculture, Ecosystems & Environment*, 136(3), 343–350. <https://doi.org/10.1016/j>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008. https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- 565 • Wang, C., Amon, B., Schulz, K., & Mehdi, B. (2021). Factors that influence nitrous oxide emissions from agricultural soils as well as their representation in simulation models: A review. *Agronomy*, 11, 770. <https://doi.org/10.3390/agronomy11040770>
- Xing, H., Wang, E., Smith, C. J., Rolston, D., & Yu, Q. (2011). Modelling nitrous oxide and carbon dioxide emission from soil in an incubation experiment. *Geoderma*, 167–168, 328–339. <https://doi.org/10.1016/j.geoderma.2011.07.003>



570

- Zhang, J., Zhu, Y., Zhang, X., Ye, M., & Yang, J. (2018). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of Hydrology*, 561, 918–929. <https://doi.org/10.1016/j.jhydrol.2018.04.065>
- Zimmermann, J., Ramsey, R., Forrester, P., Harty, M., Lanigan, G., Richards, K., et al. (2018). Assessing the performance of three frequently used biogeochemical models when simulating N₂O emissions from a range of soil types and fertiliser treatments. *Geoderma*, 331, 53–69. <https://doi.org/10.1016/j.geoderma.2018.06.004>